Manav Mittal, 202153 8

Date ....

Derivation of Normal equation

Let us assume that we have n training examples in our data set.

And let $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$ and $X = \begin{bmatrix} x_1 & x_2 & \cdots & x_m \\ & & & \\ x_1 & x_2 & \cdots & x_m \end{bmatrix}$

and weights are $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix}$

$$\text{Error} = \begin{bmatrix} y_1 - (x_1\theta_1 + x_2\theta_2 \cdots x_m\theta_m) \\ \vdots \\ \vdots \\ y_n - (x_{n1}\theta_1 + x_{n2}\theta_2 \cdots x_{nm}\theta_m) \end{bmatrix} = Y - X\theta$$

Let e_i denote the i^{th} entry of the Error matrix. and we want to minimize $e_1^2 + e_2^2 \cdots e_n^2$.

$$\text{Let } L = (\text{Error})^2$$

$$L = (Y - X\theta)^T (Y - X\theta)$$

$$= (Y^T - \theta^T X^T)(Y - X\theta) \quad (\because (AB)^T = B^T A^T)$$

$$= Y^T Y - Y^T X\theta - \theta^T X^T Y - \theta^T X^T X\theta$$

$$\frac{\partial L}{\partial \theta} = 0 - Y^T X - X^T Y - (X^T X\theta + (\theta^T X^T X)^T) = 0$$

$$= 2X^T X\theta - 2X^T Y = 0$$

$$X^T X\theta = X^T Y$$

$$\boxed{\theta = (X^T X)^{-1} (X^T Y)}$$

This is the expression of the normal equation for linear regression

Spiral